Statistica Sinica: Preprint
doi:10.5705/ss.2011.271

OPTIMAL DESIGNS FOR TWO-PARAMETER NONLINEAR MODELS WITH APPLICATION TO SURVIVAL MODELS

Maria Konstantinou, Stefanie Biedermann and Alan Kimber

Statistical Sciences Research Institute, University of Southampton

Abstract: Censoring may occur in many industrial or biomedical 'time to event' experiments. Efficient designs for such experiments are needed but finding such designs can be problematic since the statistical models involved will usually be nonlinear, making the optimal choice of design parameter dependent. We provide analytical characterisations of locally D- and c-optimal designs for a large class of models. Our results are illustrated using the natural proportional hazards parameterisation of the exponential regression model, thus reducing the numerical effort for design search substantially. We also determine designs based on standardised optimality criteria when a range of parameter values is provided by the experimenter. Different censoring mechanisms are incorporated and the robustness of designs against parameter misspecification is assessed. We demonstrate that, unlike traditional designs, the designs found perform well across a broad range of scenarios.

Key words and phrases: c-optimality, D-optimality, proportional hazards, survival analysis.

1. Introduction

Optimal designs for nonlinear models have received much attention in the literature but there is little research on designs for models with potentially censored data. Ford, Torsney and Wu (1992) consider optimal designs for nonlinear models where the response variable is distributed as a member of the exponential family and Sebastiani and Settimi (1997) prove the optimality of these designs for a logistic regression model. Sitter and Torsney (1992) study D-optimal designs for generalised linear models with multiple design variables using the geometry of the design space as in Ford, Torsney and Wu (1992), and Sitter and Torsney (1995) consider both D- and c-optimal designs for binary response models with two design variables. However neither of these papers considers the case where the data are subject to censoring. Optimal adaptive designs for the survival Koziol-Green

Statistica Sinica: Preprint
doi:10.5705/ss.2011.271

model with two treatments are discussed in Bandyopadhyay, Biswas and Bhattacharya (2010). Xu (2009) considers designs for generalised linear models with possibly censored observations which are robust with respect to model misspecifications within a certain class. For recent results on accelerated life testing see, for example, Wu, Lin and Chen (2006) and McGree and Eccleston (2010).

Proportional hazards models are considered by Becker, McDonald and Khoo (1989) who find D-optimal designs for models with one or two parameters and completely specified baseline hazard. They use geometric arguments and empirical values for the hazard rate to investigate how censoring affects the D-optimal designs for different shapes of the design region. López-Fidalgo, Rivas-López and Del Campo (2009) propose an algorithm to find D-optimal designs conditional on arrival time, where the design space consists of two treatments. They consider a two-parameter exponential regression model that requires constraints on the parameters.

Our research was motivated by the following problem. Let $T_1, \ldots, T_n$ be independent random variables indicating the survival times of the n subjects in the experiment with $t_1, \ldots, t_n$ the corresponding observed values. Also let α and β be the unknown model parameters requiring estimation and let $x_j \in \mathcal{X}$ be the experimental condition at which the jth observation is taken. In what follows, the design space $\mathcal{X}$ can be either binary, that is $\mathcal{X} = \{0, 1\}$, corresponding, for example, to two different treatments, or an interval, that is $\mathcal{X} = [u, v]$, corresponding, for example, to the doses of a drug.

The period of the experiment will be defined to be the interval $[0, c]$. We consider two different types of censoring. Type I censoring corresponds to the situation where all subjects enter the study at the same time and stay until time c or until failure, whichever is earlier. Survival times greater than c are therefore right-censored. Another scenario, relevant for many clinical trials, is random censoring. The jth individual enters the study at a random time in $[0, c]$ which is independent of the survival time. Therefore the censoring time for this individual is also random. The example we shall use to illustrate our general results is the exponential regression model in its proportional hazards parameterisation which is specified by the probability density function $f(t_j)$ with corresponding survivor

Statistica Sinica: Preprint
doi:10.5705/ss.2011.271

function $S(t_j)$,

$$f(t_j) = e^{\alpha+\beta x_j} e^{-t_j e^{\alpha+\beta x_j}}, \quad S(t_j) = e^{-t_j e^{\alpha+\beta x_j}}, \qquad (t_j > 0) \tag{1.1}$$

The parameterisation in (1.1), avoids the need to specify constraints on the parameters.

Optimal design is concerned with finding the experimental conditions at which measurements should be taken in order to draw the most precise conclusions from the data. In what follows, we consider approximate designs of the form

$$\xi = \begin{Bmatrix} x_1 & \dots & x_m \\ \omega_1 & \dots & \omega_m \end{Bmatrix},$$

where the support points x_i, $i = 1, \dots, m$, $m \leq n$ are the distinct experimental conditions in the design and the weights ω_i represent the proportion of observations in the corresponding support point. We note that $0 < \omega_i \leq 1$ and $\sum_{i=1}^m \omega_i = 1$, so a design ξ is a probability measure on the design space $\mathcal{X}$. The $n\omega_i$ are not necessarily integers for all i, so before an approximate design can be run a rounding procedure (see, for example, Pukelsheim and Rieder (1992)) must be applied. Approximate designs have the advantage of avoiding discrete optimisation during design search.

If the aim of the experiment is efficient estimation of the model parameters or functions of them, optimal designs are found by optimising a statistically meaningful functional of, for example, the Fisher information with respect to the design. Such a functional, which is called an optimality criterion, maps nonnegative definite square matrices to the real axis, and so makes matrices comparable. Typical optimality criteria are D- and c-optimality, which will be introduced in sections 2 and 3, respectively.

A recent trend in the optimal design literature is to solve problems in more generality. Hedayat, Zhong and Nie (2004) characterise D-optimal designs for a large class of two parameter models under extra assumptions on the Fisher information. However, these assumptions are not generally satisfied. In particular, these results are not applicable to model (1.1). Yang and Stufken (2009) consider Loewner optimality and an even more general class of models. They obtain a series of excellent results, showing that under some conditions, for each given design there is always a design from a simple class which is better in the

Statistica Sinica: Preprint
doi:10.5705/ss.2011.271

Loewner sense. These results were then generalised to models with more than two parameters by Yang (2010) and Dette and Melas (2011). Depending on the model, however, these conditions can be difficult to verify even using symbolic computational software.

The aim of this paper is to provide characterisations of D- and c-optimal designs under assumptions which are less restrictive and easier to verify than those in Yang and Stufken (2009) and which are satisfied by a large class of models, including model (1.1). In section 2 we develop this approach for D-optimality. Section 3 contains the corresponding results for c-optimality in the situation where only the slope parameter β is of interest. The results are illustrated by application to model (1.1) with Type I and random censoring in section 4. Since the model is nonlinear in the parameters the optimal designs found depend on the unknown model parameters, that is they are locally optimal (Chernoff (1953)). Section 5 provides analytical characterisations of the standardised maximin D- and c-optimal designs when a parameter space can be specified even in situations where the locally optimal designs are not available in closed form. In section 6, we assess the robustness of locally optimal and parameter robust designs for model (1.1) and compare their efficiency with traditional designs currently in use. A brief discussion is given in section 7. The more technical proofs can be found in the appendix.

2. D-optimal designs

A D-optimal design maximises the determinant of the Fisher information $M(\xi, \alpha, \beta)$ with respect to the design. That is, a design ξ^* is D-optimal if

$$\xi^* = \arg\max_{\xi} |M(\xi, \alpha, \beta)|.$$

A D-optimal design minimises the volume of the confidence ellipsoid for the parameter estimators, so makes the estimators as precise as possible.

Hedayat, Zhong and Nie (2004) consider two parameter models with Fisher information matrix proportional to

$$M(\xi, \alpha, \beta) = \sum_{i=1}^{m} \omega_i I(x_i, \alpha, \beta) = \sum_{i=1}^{m} \omega_i Q(\theta_i) \begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix}, \tag{2.1}$$

where $I(x_i, \alpha, \beta)$ is the Fisher information at the point x_i and $\theta_i = \alpha + \beta x_i$. Many models share this form of information matrix, including generalised linear

doi:10.5705/ss.2011.271

models (Ford, Torsney and Wu (1992), Sebastiani and Settimi (1997)), some linear heteroscedastic models and model (1.1) for different censoring schemes. Key assumptions in Hedayat, Zhong and Nie (2004) are that $Q(\theta)$ has exactly one stationary point and $\lim_{\theta\to\infty} Q(\theta)\theta^2$ is bounded. For many relevant situations however, $Q(\theta)$ is strictly increasing and $\lim_{\theta\to\infty} Q(\theta)\theta^2$ unbounded; see section 4 for example.

In this section, we concentrate on models with Fisher information of the form (2.1), which satisfy the following conditions (a)-(d). Following Ford, Torsney and Wu (1992), an equivalent problem to maximising $|M(\xi,\alpha,\beta)|$ is to maximise the determinant of this matrix with x_i replaced by θ_i, $i = 1,\ldots,m$, which will also be denoted $M(\xi,\alpha,\beta)$ in what follows. Instead of the design space $\mathcal{X}$ we will work on the transformed design space $\Theta = \alpha + \beta\mathcal{X}$ where $\beta \neq 0$. The parameter dependence of the design problem thus enters only via the transformed design space. The assumptions are therefore given for $\theta \in \mathbb{R}$, so they are valid for all possible ranges for Θ.

(a) The function $Q(\theta)$ implicitly defined in (2.1) is positive for all $\theta \in \mathbb{R}$ and twice continuously differentiable.

(b) The function $Q(\theta)$ is strictly increasing on $\mathbb{R}$.

(c) The second derivative $g''(\theta)$ of the function $g(\theta) = 2/Q(\theta)$ is an injective function.

(d) For fixed $s \in \mathbb{R}$, the function $r(\theta) = Q(\theta)(s-\theta)^2$ satisfies $r'(\theta) = 0$ for exactly two values of $\theta \in (-\infty, s]$.

We note that if assumptions (a) and (b) are satisfied, then assumption (d1): "$Q''(\theta) < 0$ for all $\theta \in \mathbb{R}$" implies assumption (d). It is thus stronger than (d), but may be easier to check in practice. For the case of c-optimality we require the extra condition

(d2) : The function $\log Q(\theta)$ is concave for $\theta \in \mathbb{R}$.

which also implies assumption (d) given that (a) and (b) are satisfied.

To allow estimation of both parameters, a design must have at least two support points. For the binary design space $\mathcal{X} = \{0,1\}$ this means that both points, 0 and 1, are support points of the D-optimal design. From Lemma 5.1.3 in Silvey (1980), it follows that for any model with Fisher information of the form (2.1) and as many points as there are model parameters the D-optimal design has equal weights. Therefore the D-optimal design ξ^* for design space $\mathcal{X} = \{0, 1\}$ is

$$\xi^* = \begin{Bmatrix} 0 & 1 \\ 0.5 & 0.5 \end{Bmatrix}.$$

For the remaining part of this section we will consider design spaces that are intervals $\mathcal{X} = [u, v]$. For a continuous explanatory variable x, D-optimal designs are invariant with respect to the design space (see, for example, Silvey (1980)). We can therefore without loss of generality consider the design space $\mathcal{X} = [0, 1]$. A useful tool for characterising D-optimal designs as well as checking the D-optimality of a candidate design is the equivalence theorem (see, for example, Silvey (1980)).

Theorem 1. *A design ξ^* is D-optimal for a model with information matrix (2.1) if the inequality*

$$d(\xi^*, \alpha, \beta) = tr\{M^{-1}(\xi^*, \alpha, \beta) I(x, \alpha, \beta)\} \leq 2,$$

holds for all $x \in [0, 1]$, with equality in the support points of ξ^*.

From Caratheodory's Theorem (see, for example, Silvey (1980)), there exists a D-optimal design with at most three support points. Lemma 1 shows that this number can be further reduced.

Lemma 1. *Let $\beta \neq 0$ and assumptions (a)-(c) be satisfied. Then the D-optimal design for a model with Fisher information (2.1) is unique and has two equally weighted support points.*

The proof of Lemma 1 modifies an idea of Biedermann, Dette and Zhu (2006), and is given in the appendix. Dette, Melas and Wong (2006) showed that approximate locally D-optimal designs for exponential regression models that can be written as a sum of exponential terms are minimally supported. Lemma 1 comes to an agreement with these results as the special case of one exponential term. We are now ready to present the main result of this section, an analytical characterisation of D-optimal designs.

Theorem 2. *Let assumptions (a)-(d) be satisfied.*
(a) If $\beta > 0$, the design

$$\xi^* = \begin{Bmatrix} x_1^* & 1 \\ 0.5 & 0.5 \end{Bmatrix},$$

is D-optimal on $\mathcal{X}$, where $x_1^* = 0$ if $\beta < 2Q(\alpha)/Q'(\alpha)$. Otherwise x_1^* is the unique solution of the equation $\beta(x_1 - 1) + 2Q(\alpha + \beta x_1)/Q'(\alpha + \beta x_1) = 0$.

(b) If $\beta < 0$, the design

$$\xi^* = \begin{Bmatrix} 0 & x_2^* \\ 0.5 & 0.5 \end{Bmatrix},$$

is D-optimal on $\mathcal{X}$, where $x_2^* = 1$ if $\beta > -2Q(\alpha + \beta)/Q'(\alpha + \beta)$. Otherwise x_2^* is the unique solution of the equation $\beta x_2 + 2Q(\alpha + \beta x_2)/Q'(\alpha + \beta x_2) = 0$.

Theorem 2, the proof of which is in the appendix, provides a complete classification of D-optimal designs. Depending on some easily verifiable conditions on the parameters, the design problem has either been reduced to an optimisation problem in one variable or been solved entirely.

3. c-optimal designs

For model (1.1), in the case of a binary design space, the parameter β represents the effect on the hazard of death of, for example, the new treatment whereas for an interval design space, β represents, for example, the effect of increasing the dose of a drug. Therefore, often interest centers on estimating the parameter β while treating α as a nuisance parameter. In this situation, an appropriate optimality criterion is c-optimality for β which minimises the asymptotic variance of the maximum likelihood estimator $\hat{\beta}$. Thus a design ξ^* is c-optimal for β if the vector $(0 \;\; 1)^T$ is in the range of $M(\xi^*, \alpha, \beta)$ and

$$\xi^* = \arg\min_{\xi} (0 \quad 1)\, M^-(\xi, \alpha, \beta) \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{3.1}$$

where M^- is a generalised inverse of the matrix M.
Lemma 2, which is proved in the appendix shows that the c-optimal design for β is supported on two points for all censoring structures.

Lemma 2. *For any choice of α, $\beta \neq 0$ and any model with Fisher information (2.1) there exists a c-optimal design for estimating β with exactly two support points.*

Statistica Sinica: Preprint
doi:10.5705/ss.2011.271

From Pukelsheim and Torsney (1991), we obtain an expression for the optimal weights, for example, a c-optimal design ξ^* for β with support points x_1^* and x_2^* is given by

$$\xi^* = \left\{ \begin{array}{cc} x_1^* & x_2^* \\ \frac{\sqrt{Q(\alpha+\beta x_2^*)}}{\sqrt{Q(\alpha+\beta x_1^*)}+\sqrt{Q(\alpha+\beta x_2^*)}} & \frac{\sqrt{Q(\alpha+\beta x_1^*)}}{\sqrt{Q(\alpha+\beta x_1^*)}+\sqrt{Q(\alpha+\beta x_2^*)}} \end{array} \right\}. \quad (3.2)$$

The design problem for the binary design space $\mathcal{X} = \{0, 1\}$ has thus been solved completely. It remains to find the optimal support points for an interval design space $\mathcal{X} = [u, v] \subset \mathbb{R}$. An analytical characterisation of the c-optimal designs for β on the interval $\mathcal{X} = [u, v]$ for models with information matrix of the form (2.1) is given in Theorem 3, which is proved in the appendix.

Theorem 3. *Let assumptions (a)-(c) and (d2) be satisfied.*
(a) If $\beta > 0$, the design ξ^* with support points x_1^* and v and the optimal weights given in (3.2) is c-optimal for β, where $x_1^* = u$ if

$$\beta(u - v) + 2Q(\alpha + \beta u)/Q'(\alpha + \beta u) \left(1 + \sqrt{Q(\alpha + \beta u)}/\sqrt{Q(\alpha + \beta v)}\right) > 0. \quad (3.3)$$

Otherwise x_1^* is the unique solution of the equation

$$\beta(x_1 - v) + 2Q(\alpha + \beta x_1)/Q'(\alpha + \beta x_1) \left(1 + \sqrt{Q(\alpha + \beta x_1)}/\sqrt{Q(\alpha + \beta v)}\right) = 0. \quad (3.4)$$

(b) If $\beta < 0$, the design ξ^* with support points u and x_2^* and the optimal weights given in (3.2) is c-optimal for β, where $x_2^* = v$ if

$$\beta(u - v) - 2Q(\alpha + \beta v)/Q'(\alpha + \beta v) \left(1 + \sqrt{Q(\alpha + \beta v)}/\sqrt{Q(\alpha + \beta u)}\right) < 0.$$

Otherwise x_2^* is the unique solution of the equation

$$\beta(u - x_2) - 2Q(\alpha + \beta x_2)/Q'(\alpha + \beta x_2) \left(1 + \sqrt{Q(\alpha + \beta x_2)}/\sqrt{Q(\alpha + \beta u)}\right) = 0.$$

4. Application to an exponential regression model

In this section we apply the previous results to model (1.1) for an interval design space. We briefly discuss the special case of no censoring, corresponding to $c = \infty$, a study running for as long as necessary to record all survival times. From (1.1), the log-likelihood at x_j is

$$l(\alpha, \beta, x_j) = \alpha + \beta x_j - t_j e^{\alpha+\beta x_j},$$

Statistica Sinica: Preprint
doi:10.5705/ss.2011.271

and thus the Fisher information at the point x_j is

$$I(x_j, \alpha, \beta) = \begin{pmatrix} 1 & x_j \\ x_j & x_j^2 \end{pmatrix},$$

since $E(T_j) = 1/e^{\alpha+\beta x_j}$. We note that in this case the Fisher information is in fact the same as for the linear model $Y_j = \alpha + \beta x_j + \epsilon_j$ for independent identically distributed errors $\epsilon_j \sim \mathcal{N}(0, \sigma^2)$. It is well known (see, for example, Atkinson, Donev and Tobias (2007)) that the D-optimal design for this model is equally supported at the end points of the design space $\mathcal{X}$. For the c-optimality case we observe that the objective function defined in (3.1) for model (1.1) is the inverse of the determinant of the Fisher information. Therefore the D-optimal design allocating equal weight to the end points of the design space $\mathcal{X}$ is also c-optimal for β in this case. We now consider two different types of censoring.

4.1. Type I censoring

Type I censoring means that the censoring time c is fixed and common for all individuals. This occurs, for example, when all individuals have been recruited at the same time to a study of duration c. If failure has not occurred by the end of the study the observation is right-censored. Assume without loss of generality that the first k observations are failure times. The log-likelihood for model (1.1), $l(\alpha, \beta, x_j)$, is then given by

$$l(\alpha, \beta, x_j) = \log\{\prod_{j=1}^{k} f(t_j) \prod_{j=k+1}^{n} S(t_j)\} = \sum_{j=1}^{k}(\alpha + \beta x_j) - \sum_{j=1}^{n} t_j e^{\alpha+\beta x_j}.$$

Now the expected survival time at x_j is

$$E(T_j) = \int_0^c t e^{\alpha+\beta x_j} e^{-te^{\alpha+\beta x_j}}\, dt + cP(T_j > c) = (1 - e^{-ce^{\alpha+\beta x_j}})/e^{\alpha+\beta x_j},$$

and hence the Fisher information at x_j is

$$I(x_j, \alpha, \beta) = (1 - e^{-ce^{\alpha+\beta x_j}}) \begin{pmatrix} 1 & x_j \\ x_j & x_j^2 \end{pmatrix} = q(\theta_j) \begin{pmatrix} 1 & x_j \\ x_j & x_j^2 \end{pmatrix},$$

which is of the form (2.1). It can be verified that assumptions (a)-(d) and (d2) are true for $Q(\theta) = q(\theta)$ and hence Theorems 2 and 3 for D and c-optimal designs respectively, hold for the case of Type I censoring.

Statistica Sinica: Preprint
doi:10.5705/ss.2011.271

4.2. Random censoring

Random censoring occurs, for example, if the jth individual enters the study at random time $Z_j \in [0, c]$, where Z_j is independent of the survival time T_j. Now the censoring time $C_j = c - Z_j$ for this individual is also random. In what follows we assume that the censoring times $C_1, \ldots, C_n$ follow a continuous uniform distribution on $[0, c]$. The probability density function of C_j is therefore $f_c(c_j) = 1/c$, and the conditional density of T_j given $C_j = c_j$ is given in (1.1). The joint density of C_j and T_j is the product of $f(t_j|c_j)$ and $f_c(c_j)$, and thus the log-likelihood at x_j is

$$l(\alpha, \beta, x_j) = -\log c + \alpha + \beta x_j - t_j e^{\alpha+\beta x_j}.$$

Using the law of iterated expectations,we find that

$$\begin{aligned} E(T_j) = E(E(T_j|C_j = c_j)) &= \int_0^c (1 - e^{-c_j e^{\alpha+\beta x_j}})/ce^{\alpha+\beta x_j}\, dc_j \\ &= \left(ce^{\alpha+\beta x_j} + e^{-ce^{\alpha+\beta x_j}} - 1\right)/ce^{2(\alpha+\beta x_j)}, \end{aligned}$$

and so the Fisher information at point x_j is

$$I(x_j, \alpha, \beta) = \left(ce^{\alpha+\beta x_j} + e^{-ce^{\alpha+\beta x_j}} - 1\right)/ce^{\alpha+\beta x_j} \begin{pmatrix} 1 & x_j \\ x_j & x_j^2 \end{pmatrix} = \tilde{q}(\theta_j) \begin{pmatrix} 1 & x_j \\ x_j & x_j^2 \end{pmatrix}.$$

Again this is of the form (2.1) and it can be shown that assumptions (a)-(d) and (d2) hold for $Q(\theta) = \tilde{q}(\theta)$.

For positive slope parameter β the functions $q(\alpha + \beta x)$ and $\tilde{q}(\alpha + \beta x)$ are increasing with x, whereas they are decreasing if β is negative. Therefore from (3.2) the c-optimal weight corresponding to the smaller support point is greater than the other weight if $\beta > 0$ and smaller if $\beta < 0$. This means, for example, that more patients are allocated to the more effective dose. It is interesting to note that the popular equal allocation rule leads to a suboptimal design.

5. Standardised optimal designs

The optimal designs found above depend on the model parameters which are unknown in practice. Nevertheless, in many practical situations some information about the parameter values can be provided by the experimenter. For example, parameter α may represent the baseline hazard for a standard treatment and hence even precise knowledge of its value might be available, whereas

Statistica Sinica: Preprint
doi:10.5705/ss.2011.271

for the parameter β the experimenter can specify a range of values for a clinically significant improvement with new treatment. We further assume that the experimenter has no preference for specific β-values and that the total duration of the study, c, is known.

Following Dette (1997) we seek for designs that maximise the worst efficiencies with respect to the locally optimal designs over a range of parameter values. This allows us to construct robust designs in the sense that they protect against the worst case scenario. Dette and Sahm (1998) consider both the standardised maximum variance criterion as well as Elfving's maximum variance optimality criterion for nonlinear models and showed that in some cases the optimal designs based on the latter criterion may become inefficient. A design ξ^* maximising the criterion

$$\Phi(\xi) = \min\left\{ \frac{|M(\xi,\alpha,\beta)|}{|M(\xi^*_\beta,\alpha,\beta)|} \;\middle|\; \beta \in [\beta_0, \beta_1] \right\}, \tag{5.1}$$

is called a standardised maximin D-optimal design and a design ξ^* maximising the criterion

$$\Phi(\xi) = \min\left\{ \frac{(0 \quad 1) M^-(\xi^*_\beta,\alpha,\beta)\binom{0}{1}}{(0 \quad 1) M^-(\xi,\alpha,\beta)\binom{0}{1}} \;\middle|\; \beta \in [\beta_0, \beta_1] \right\}, \tag{5.2}$$

is called a standardised maximin c-optimal design for β, where ξ^*_β is the locally optimal design. Criteria (5.1) and (5.2) seek a design that maximises the worst D-efficiency and c-efficiency respectively, given by

$$eff_D(\xi) = \left(\frac{|M(\xi,\alpha,\beta)|}{|M(\xi^*_\beta,\alpha,\beta)|} \right)^{\frac{1}{2}}, \tag{5.3}$$

and

$$eff_c(\xi) = \frac{(0 \quad 1)\, M^-(\xi^*_\beta,\alpha,\beta)\binom{0}{1}}{(0 \quad 1)\, M^-(\xi,\alpha,\beta)\binom{0}{1}}. \tag{5.4}$$

In the case of a binary design space the locally D-optimal design is equally supported at 0 and 1 for any parameter values, so no further investigation needs to be done. For an interval design space $\mathcal{X} = [0, 1]$, the following theorem provides an analytical characterisation of the standardised maximin D-optimal two point design for a given range of negative β-values and its proof is given in the appendix.

Theorem 4. *Let $\beta \in [\beta_0, \beta_1]$ where $\beta_1 < 0$, α be fixed and assumptions (a), (b) and (d2) be satisfied. The standardised maximin D-optimal two-point design is equally supported at points 0 and x_2^* where $x_2^* = 1$ if $\beta_0 > -2Q(\alpha+\beta_0)/Q'(\alpha+\beta_0)$. Otherwise x_2^* is the solution of the equation*

$$Q(\alpha+\beta_0 x)Q(\alpha+\beta_1 x_{\beta_1})x_{\beta_1}^2 = Q(\alpha+\beta_1 x)Q(\alpha+\beta_0 x_{\beta_0})x_{\beta_0}^2, \qquad (5.5)$$

and x_{β_0}, x_{β_1} are the solutions of the equation $\beta x + 2Q(\alpha+\beta x)/Q'(\alpha+\beta x) = 0$, $0 < x \leq 1$ for β_0 and β_1 respectively.

Note that Theorem 4 applies when $\beta < 0$. The proof used in this case does not work when $\beta > 0$ and this is a topic for further investigation.

As shown in section 3, the locally c-optimal design for β for the case of a binary design space depends on the model parameters. Theorem 5 gives an analytical characterisation of the standardised maximin c-optimal design for β, in the case of a binary design space and is also proven in the appendix.

Theorem 5. *Let $\beta \in [\beta_0, \beta_1]$, α be fixed and assumptions (a), (b) and (d2) be satisfied. Also let the design space to be binary, that is $\mathcal{X} = \{0, 1\}$. The standardised maximin c-optimal two-point design is*

$$\xi^* = \begin{Bmatrix} 0 & 1 \\ \omega^* & 1-\omega^* \end{Bmatrix},$$

where $\omega^* = (\omega_{\beta_0} + \omega_{\beta_1})/2$ and ω_{β_0} and ω_{β_1} is the value of optimal weight on the smaller support point for the locally c-optimal design for β given in (3.2) for β_0 and β_1 respectively.

6. Robustness analysis

In the following we assess the robustness of our designs by calculating their efficiency if the parameters have been misspecified. As a starting point we used the maximum likelihood estimates for α and β from the Freireich data (Freireich et.al. (1963)) -2.163 and -1.526 respectively. In the efficiency calculations we used $c = 30$.

To compare the performance of an arbitrary design ξ to a locally D-optimal design ξ^* we use the D-efficiency (5.3), whereas for the comparison of ξ to a locally c-optimal design ξ^* we use the c-efficiency (5.4). Type I censoring is assumed throughout this numerical example for demonstration purposes.

Statistica Sinica: Preprint
doi:10.5705/ss.2011.271

6.1. Locally D-optimal designs

Here we consider locally D-optimal designs for a vector of parameter values $\boldsymbol{\gamma} = (\alpha, \beta)$. The value of the maximum likelihood estimator for α was used, whereas the β-values were chosen so that we have small, medium and large treatment effect. Table 6.1 gives the parameter vectors used and the corresponding D-efficiencies of the locally D-optimal designs when the parameter values are misspecified.

Table 6.1: D-efficiencies for some selected locally D-optimal designs

	Design			
Parameter vector	ξ_{γ_0}	ξ_{γ_1}	ξ_{γ_2}	ξ_{γ_3}
$\boldsymbol{\gamma_0} = (-2.163, -0.1)$	1	1	1	0.900
$\boldsymbol{\gamma_1} = (-2.163, -0.405)$	1	1	1	0.905
$\boldsymbol{\gamma_2} = (-2.163, -1.526)$	1	1	1	0.946
$\boldsymbol{\gamma_3} = (-2.163, -2.623)$	0.992	0.992	0.992	1

For the first three sets of parameter values the locally D-optimal design is the "standard design" supported at 0 and 1, with equal weights whereas ξ_{γ_3} is equally supported at 0 and 0.9. The "standard design" has very high D-efficiency for all the values of the parameter vectors. The lowest efficiency, 0.900, is obtained if the true value is $\boldsymbol{\gamma_0}$ and the experimenter has misspecified this value as $\boldsymbol{\gamma_3}$ and hence used the design ξ_{γ_3}. In other words if the experimenter has used design ξ_{γ_3} assuming a large treatment effect when the true effect is actually very small, the D-efficiency is 0.9 which is quite satisfying. Hence ξ_{γ_3} seems to be a good alternative to the "standard design" if, for example, the experimenter does not want to expose the patients at the higher drug dose.

6.2. Locally c-optimal designs

For the same vectors of parameter values used in section 6.1, the support points of the locally c-optimal designs are always 0 and 1. The c-optimal weights were found using (3.2) and are shown in Table 6.2. The c-efficiencies of each of the above designs was also calculated when the parameter values are misspecified and are presented in Table 6.3.

Generally the locally c-optimal designs have high c-efficiencies for all the four sets of parameter values. The lowest efficiency, 0.8772, is obtained when the

Table 6.2: Weights for some selected locally c-optimal designs

	Design			
Weight	ξ_{γ_0}	ξ_{γ_1}	ξ_{γ_2}	ξ_{γ_3}
ω_1	0.498	0.491	0.425	0.323
ω_2	0.502	0.509	0.575	0.677

Table 6.3: c-effficiencies for the locally c-optimal designs of Table 2

	Design			
Parameter vector	ξ_{γ_0}	ξ_{γ_1}	ξ_{γ_2}	ξ_{γ_3}
$\boldsymbol{\gamma_0}$	1	0.9998	0.9782	0.8772
$\boldsymbol{\gamma_1}$	0.9998	1	0.9824	0.8864
$\boldsymbol{\gamma_2}$	0.9787	0.9828	1	0.9552
$\boldsymbol{\gamma_3}$	0.8908	0.8991	0.9597	1

assumed value is $\boldsymbol{\gamma_3}$ and the true value is $\boldsymbol{\gamma_0}$. We also observe that the design ξ_{γ_2}, which is locally c-optimal for the parameter values towards the center of the parameter space, has a lowest efficiency of 0.9597 and hence is more robust than the other three designs.

6.3. Standardised maximin optimal designs

According to the analysis in section 5 we can find the standardised maximin D- and c-optimal designs for the range of β-values used above which are denoted by ξ_{γ_4} in both cases. We note that although here we consider the case of an interval design space all the locally c-optimal designs found in section 6.2 are supported at points 0 and 1 and so the result of theorem 5 can be used.

The standardised maximin D-optimal design is supported at 0 and 0.993, with equal weights and is locally D-optimal for $\boldsymbol{\gamma_4} = (-2.163, -2.380)$, whereas the standardised maximin c-optimal design allocates $41,1\%$ of the observations at the experimental point 0 and the rest at point 1, and is locally c-optimal for $\boldsymbol{\gamma_4} = (-2.163, -1.690)$. The minimum (median) efficiencies are 0.993 (0.993) for the D-optimal design and 0.969 (0.974) for the c-optimal design. For both of the above designs the minimum efficiencies are obtained at $\boldsymbol{\gamma_0}$ and $\boldsymbol{\gamma_3}$.

6.4. Cluster designs

This is a modification of the method introduced by Dror and Steinberg (2006). For each one of 1000 parameter vectors, found by drawing 1000 β-values from a uniform distribution on the interval from -2.623 to -0.1, the locally D- and c-optimal designs were obtained. Then a clustering algorithm was applied for both D- and c-optimal designs and the resulting cluster designs were allowed to have unequally weighted support points (Biedermann and Woods (2011)). The cluster centroids are chosen as the support points and each weight is chosen to be proportional to the corresponding cluster size, reflecting the relative importance of each cluster.

The number of clusters for the D-optimal designs was chosen to vary from 2 to 6 and for each value the D-efficiency of a cluster design was calculated via (5.3) relative to each of the 1000 locally D-optimal designs. The minimum and median efficiencies are found to be the same for all the cluster designs (0.993 and 0.997 respectively) and this may be a result of the very low weight that all of our cluster designs give to experimental points other than 0 and 1.

The support points of the 1000 locally c-optimal designs are always 0 and 1, hence the cluster design can only have two support points which will be the experimental points 0 and 1. Also the clustering here was applied to *design* points, rather than *support* points as the support points of the locally c-optimal points have differing weights. The resulting cluster design allocates 43% of the observations at 0 and the rest at 1, and performs well as the minimum and median efficiencies found via (5.4) are 0.956 and 0.990 respectively.

6.5. Comparison of designs

First we compare the performance of the following 11 designs: the locally D-optimal designs $\xi_{\gamma_0}, \ldots, \xi_{\gamma_3}$, the standardised maximin D-optimal design ξ_{γ_4}, the cluster designs $\xi_1, \ldots, \xi_5$ and the equally spaced design ξ_0 with support points 0, 0.5, 1 and equal weights. The D-efficiency (5.3) of each of the above designs is calculated with respect to each of the 1000 locally optimal designs and the results are summarised in Figure 6.1 following Woods, Lewis, Eccleston and Russell (2006). Design ξ_0 was omitted since it was clearly outperformed.

From Figure 6.1 we observe that the standardised maximin D-optimal design ξ_{γ_4} is indeed the one with the highest minimum efficiency but it also has a

Statistica Sinica: Preprint
doi:10.5705/ss.2011.271

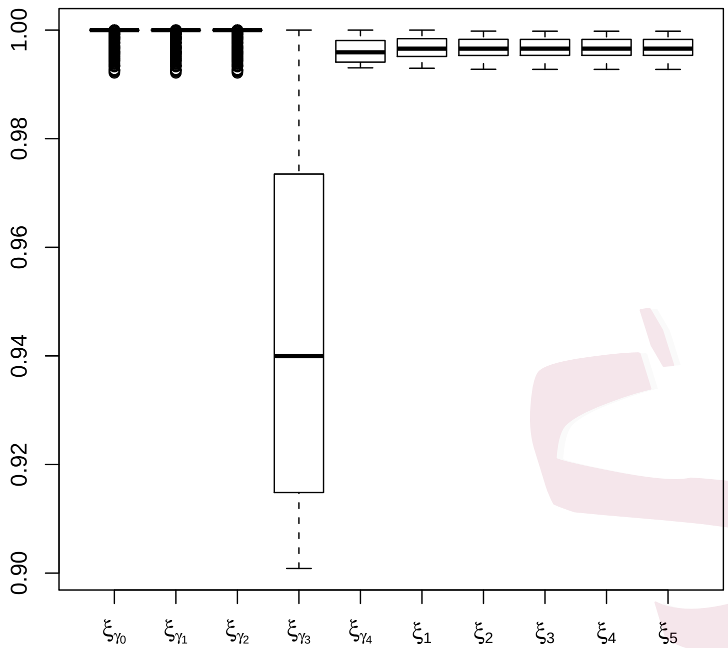

Figure 6.1: Boxplots of D-efficiencies calculated for 10 different designs for 1000 parameter vectors

lower median efficiency. Hence there is a trade off between protecting against the worst case scenario with the lowest efficiency and having a worse median efficiency. The cluster designs $\xi_2, \ldots, \xi_5$ with more than 2 support points are useful since they allow for linearity of the regression to be checked and they do not perform worse than the two-point design ξ_1. Also all five of the cluster designs are good alternatives to locally optimal designs and have similar performance to the maximin design.

The comparison of the c-optimal designs is shown in Figure 6.2. The designs used are the locally c-optimal designs $\xi_{\gamma_0}, \ldots, \xi_{\gamma_3}$, the standardised maximin c-optimal design ξ_{γ_4} and the two-point cluster design ξ_1.

Among the locally c-optimal designs $\xi_{\gamma_0}, \ldots, \xi_{\gamma_3}$, only ξ_{γ_2} performs well across the parameter space while the rest are not so good for some scenarios. As for D-optimality we can observe the trade off between best minimum efficiency and a lower median efficiency for the standardised maximin c-optimal design ξ_{γ_4}. Overall both the standardised maximin optimal design as well as the cluster design ξ_1 are good alternatives to the locally optimal designs.

Statistica Sinica: Preprint
doi:10.5705/ss.2011.271

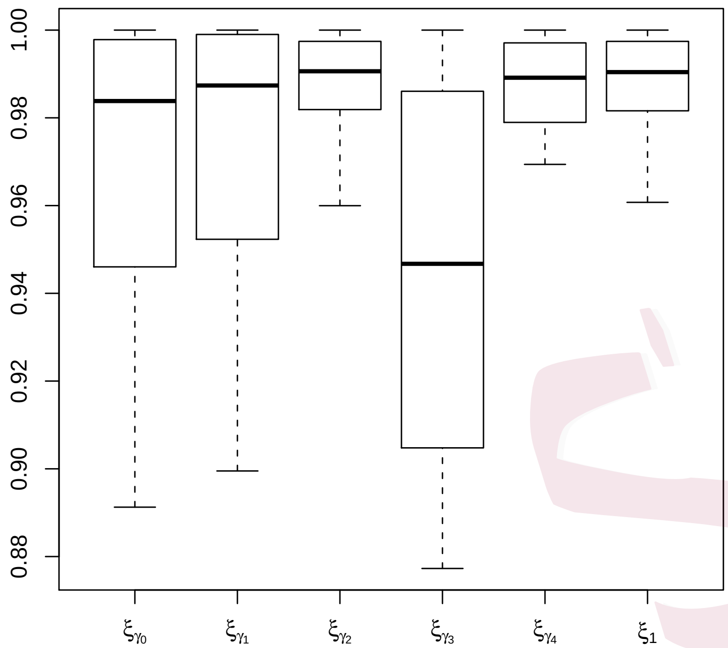

Figure 6.2: Boxplots of c-efficiencies calculated for 6 different designs for 1000 parameter vectors

7. Discussion

Time to event models used in biomedical and industrial applications are usually nonlinear and hence the optimal designs depend on the unknown model parameters. To overcome this difficulty robust designs must be constructed which will perform well across a wide range of parameter values. Another difficulty in finding optimal designs for these applications is that the data are often subject to censoring.

For models with Fisher information of the form (2.1) we have provided a complete classification of locally D- and c-optimal designs. Our assumptions are somewhat less restrictive and easier to check than those provided by Yang and Stufken (2009) for more general models. Our results were then applied to the proportional hazards parameterisation of the exponential regression model (1.1), for the cases of Type I and random censoring. For certain parameter values the optimal design is not the "standard design" supported at 0 and 1 with equal weights which is the one usually used in these experiments.

In order to construct robust designs we have found optimal designs based on

Statistica Sinica: Preprint
doi:10.5705/ss.2011.271

standardised maximin criteria when there is some knowledge about the parameter values (that is a range of values is specified), which maximise the worst efficiency among all two-point designs. To the best of our knowledge, Theorem 4 is the first analytical characterisation of standardised maximin D-optimal designs in a situation where the locally D-optimal designs are not available in closed form. Additionally, cluster designs were built from the locally optimal designs for a specific set of parameter values and their computation was facilitated by our results for the locally optimal designs. In section 6 we have shown that good alternatives to the locally optimal designs are the cluster designs which in some cases (see D-optimality), have more than 2 support points, thereby enabling the linearity of the regression function to be checked.

Appendix

Proof of Lemma 1. Let α and $\beta > 0$ be fixed and $\alpha + \beta x = \theta$. The case where $\beta < 0$ can be shown analogously and is therefore not presented. From Theorem 1, we obtain that a D-optimal design ξ^* must satisfy the inequality

$$z(\theta) := z_1 + z_2\theta + z_3\theta^2 \leq 2/Q(\theta) =: g(\theta) \quad \forall \theta \in [\alpha, \alpha + \beta],$$

for some coefficients $z_1, z_2, z_3 \in \mathbb{R}$, with equality at the support points of ξ^*.
Now suppose a D-optimal design has three support points, $\alpha \leq \theta_1 < \theta_2 < \theta_3 \leq \alpha + \beta$. Then $z(\theta_i) = g(\theta_i)$, $i = 1, 2, 3$. By Cauchy's mean value theorem, there exist points $\tilde{\theta}_i$, $i = 1, 2$ such that $\theta_1 < \tilde{\theta}_1 < \theta_2 < \tilde{\theta}_2 < \theta_3$ and $z'(\tilde{\theta}_i) = g'(\tilde{\theta}_i)$. Since $z(\theta) \leq g(\theta)$ on $[\alpha, \alpha + \beta]$, we also have $z'(\theta_2) = g'(\theta_2)$. By the mean value theorem, there exist two points $\hat{\theta}_i$, $i = 1, 2$ such that $\tilde{\theta}_1 < \hat{\theta}_1 < \theta_2 < \hat{\theta}_2 < \tilde{\theta}_2$ and $z''(\hat{\theta}_i) = g''(\hat{\theta}_i)$. Now $z''(\theta)$ is constant, so can intersect with $g''(\theta)$ at most once on $[\alpha, \alpha + \beta]$, which contradicts the assumption of three (or more) support points. Hence a D-optimal design has exactly two support points, and so the weights must be equal.
Let ξ_1 and ξ_2 be two D-optimal designs. By log-concavity of the D-criterion, the design $\xi_3 = 0.5\xi_1 + 0.5\xi_2$ must also be a D-optimal. However, if ξ_1 and ξ_2 are different, ξ_3 has more than two support points, which contradicts the result above. Hence the D-optimal design is unique.

□

Proof of Theorem 2. We give only a sketch of the proof of part (a), omitting tedious calculations. The proof of part (b) follows along the same lines using

Statistica Sinica: Preprint
doi:10.5705/ss.2011.271

symmetry arguments and is therefore omitted.

Let $\beta > 0$. A careful inspection of the determinant of the Fisher information (2.1) for a design with two support points $x_1, x_2 \in [0,1]$, with $x_1 < x_2$, yields that regardless of the value of x_1, the determinant is increasing with x_2 and therefore maximised for $x_2 = 1$. It therefore remains to maximise the function

$$r(\alpha + \beta x_1) = Q(\alpha + \beta x_1)(x_1 - 1)^2, \quad 0 \le x_1 < 1.$$

Using assumption (d), $r(\alpha+\beta x_1)$ has exactly two turning points on $(-\infty, 1]$, one of which is a minimum at $x_1 = 1$, hence the other one must be a maximum. If this maximum is attained outside the design space, $r(\alpha+\beta x_1)$ is maximised at $x_1 = 0$, which will then be the second support point of the D-optimal design. This occurs if and only $\frac{dr(\alpha+\beta x_1)}{dx_1} < 0$ at $x_1 = 0$, which is equivalent to $\beta < 2Q(\alpha)/Q'(\alpha)$. Otherwise the point at which the maximum is attained will be the second support point. This is found by solving $\frac{dr(\alpha+\beta x_1)}{dx_1} = 0$, which is equivalent to solving $\beta(x_1 - 1) + 2Q(\alpha + \beta x_1)/Q'(\alpha + \beta x_1) = 0$.

□

Proof of Lemma 2. From Caratheodory's theorem we obtain that there exists a c-optimal design for β with at most two support points. We now assume that there exists an optimal design $\tilde{\xi}$ with only one support point $\tilde{\theta}$ and find a contradiction. For estimability we require that $(0 \quad 1)^T$ is in the range of $M(\xi, \alpha, \beta)$, for example there exists a vector $\eta = (\eta_1, \eta_2)^T \in \mathbb{R}^2$ such that

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} = Q(\tilde{\theta}) \begin{pmatrix} 1 & \tilde{\theta} \\ \tilde{\theta} & \tilde{\theta}^2 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}. \tag{7.1}$$

From the first equation in (7.1), we obtain that $Q(\tilde{\theta})\eta_1 = -Q(\tilde{\theta})\eta_2\tilde{\theta}$. Substituting this into the second equation yields $1 = 0$, and therefore no c-optimal design for β with only one support point exists.

□

Proof of Theorem 3. We give only a sketch of the proof of part (a), omitting tedious calculations. The proof of part (b) is similar and therefore omitted.

Let $\beta > 0$. Substituting the expressions for the optimal weights from (3.2), we obtain for the objective function defined in (3.1):

$$k(x_1, x_2) := \left(1/\sqrt{Q(\alpha + \beta x_1)} + 1/\sqrt{Q(\alpha + \beta x_2)}\right)^2 / (x_1 - x_2)^2.$$

Statistica Sinica: Preprint
doi:10.5705/ss.2011.271

Holding the smaller support point x_1 fixed, $k(x_1, x_2)$ is decreasing with x_2 and therefore attains its minimum in the design space $\mathcal{X} = [u, v]$ at the upper boundary v, so v is the larger support point.

The function $k(x_1, v)$ has exactly one point where the derivative is zero on $(-\infty, v]$, that is there is at most one turning point in $[u, v]$, which is a minimum since $\lim_{x_1 \to -\infty} k(x_1, v) = \lim_{x_1 \to v} k(x_1, v) = \infty$. If this minimum is not in the design space, its lower boundary, u, is the smaller support point. This situation occurs if and only if $\frac{dk(x_1,v)}{dx_1} > 0$ at $x_1 = u$, which is equivalent to condition (3.3). Otherwise, the point at which the minimum is attained is the smaller support point and it can be found solving $\frac{dk(x_1,v)}{dx_1} = 0$, which is equivalent to solving (3.4). □

Proof of Theorem 4. Using condition (d2) the function $\beta + 2Q(\alpha+\beta)/Q'(\alpha+\beta)$ is increasing with β as the sum of two increasing functions. Hence if $\beta_0 + 2Q(\alpha + \beta_0)/Q'(\alpha + \beta_0) > 0$ then $\beta + 2Q(\alpha + \beta)/Q'(\alpha + \beta) > 0$ for all $\beta \in [\beta_0, \beta_1]$ and using part (b) in Theorem 2 the locally D-optimal design ξ^*_β is equally supported at points 0 and 1 for all $\beta \in [\beta_0, \beta_1]$. Hence the standardised maximin D-optimal design is also supported at 0 and 1 with equal weights.

Now let $\beta_0 \leq -2Q(\alpha + \beta_0)/Q'(\alpha + \beta_0)$. Since $\beta + 2Q(\alpha + \beta)/Q'(\alpha + \beta)$ is increasing with β there exists $\beta^* \in (\beta_0, \beta_1]$ such that $\beta + 2Q(\alpha+\beta)/Q'(\alpha+\beta) > 0$ for all $\beta \geq \beta^*$. Again using part (b) in Theorem 2 the locally D-optimal design ξ^*_β is equally supported at points 0 and x_β where $x_\beta = 1$ for $\beta \geq \beta^*$. Otherwise x_β is the solution of the equation

$$\beta x_\beta + 2Q(\alpha + \beta x_\beta)/Q'(\alpha + \beta x_\beta) = 0, \quad 0 < x_\beta \leq 1. \tag{7.2}$$

From (5.3) the D-efficiency of a two-point design ξ equally supported at 0 and x is given by

$$eff_D(\xi) = \left(\frac{Q(\alpha + \beta x)x^2}{Q(\alpha + \beta x_\beta)x_\beta^2} \right)^{\frac{1}{2}} := (u(x, \beta))^{\frac{1}{2}}.$$

For $\beta \geq \beta^*$, $x_\beta = 1$ and for fixed $0 < x \leq 1$

$$\frac{du(x, \beta)}{d\beta} = x^2/Q^2(\alpha + \beta) \left[Q'(\alpha + \beta x)xQ(\alpha + \beta) - Q(\alpha + \beta x)Q'(\alpha + \beta) \right],$$

which is non-positive for all $\beta \in [\beta^*, \beta_1]$ using condition (d2). Hence for fixed x, $u(x, \beta)$ is minimised at β_1.

Statistica Sinica: Preprint
doi:10.5705/ss.2011.271

For $\beta < \beta^*$ and fixed $0 < x \leq 1$, solving the equation $\frac{du(x,\beta)}{d\beta} = 0$ is equivalent to solving

$$Q'(\alpha+\beta x)xQ(\alpha+\beta x_\beta)x_\beta - Q(\alpha+\beta x)\left[Q'(\alpha+\beta x)(x_\beta + \beta\frac{dx_\beta}{d\beta})x_\beta + 2Q(\alpha+\beta x_\beta)\frac{dx_\beta}{d\beta}\right] = 0,$$

and using equation (7.2) becomes

$$\beta x + 2Q(\alpha+\beta x)/Q'(\alpha+\beta x) = 0.$$

This has a unique solution β such that $x_\beta = x$. So the function $\beta \to u(x,\beta)$ is unimodal for fixed x and it is minimised at β_0 or β_1. We note that if $\beta_1 + 2Q(\alpha + \beta_1)/Q'(\alpha+\beta_1) \leq 0$ then for all $\beta \in [\beta_0, \beta 1]$, $\beta + 2Q(\alpha+\beta)/Q'(\alpha+\beta) \leq 0$ and x_β is the solution of equation (7.2). Following the same arguments as in the above case for fixed $0 < x \leq 1$, the function $\beta \to u(x,\beta)$ is unimodal and minimised at β_0 or β_1.

Hence the standardised maximin design can be found by maximising

$$\Phi(\xi) = \min\left\{u(x,\beta_0), u(x,\beta_1)\right\}.$$

This maximisation can be divided into maximisation over the sets

$$M_< := \left\{x \in (0,1] \;\middle|\; u(x,\beta_0) < u(x,\beta_1)\right\}$$
$$M_> := \left\{x \in (0,1] \;\middle|\; u(x,\beta_0) > u(x,\beta_1)\right\}$$
$$M_= := \left\{x \in (0,1] \;\middle|\; u(x,\beta_0) = u(x,\beta_1)\right\}$$

Now assume that the standardised maximin D-otpimal design is in the set $M_<$ and so we must maximise the function $u(x,\beta_0)$. Taking its first derivative with respect to x and equating it to zero yields

$$\beta_0 x + 2Q(\alpha+\beta_0 x)/\beta Q'(\alpha+\beta_0 x) = 0 \Rightarrow x = x_{\beta_0}.$$

Hence $(u(x_{\beta_0},\beta_0))^{\frac{1}{2}} = 1 < (u(x_{\beta_0},\beta_1))^{\frac{1}{2}}$ which is a contradiction since the efficiency defined in (5.3) is always less than or equal to one. Following similar arguments for set $M_>$ also leads to a contradiction and so the standardised maximin D-optimal design can be found by solving $u(x,\beta_0) = u(x,\beta_1)$ which is equivalent to solving

$$Q(\alpha+\beta_0 x)Q(\alpha+\beta_1 x_{\beta_1})x_{\beta_1}^2 = Q(\alpha+\beta_1 x)Q(\alpha+\beta_0 x_{\beta_0})x_{\beta_0}^2.$$

Statistica Sinica: Preprint
doi:10.5705/ss.2011.271

□

Proof of Theorem 5. For the case of a binary design space it has been showed that the locally c-optimal design for estimating β allocates ω_β observations at point 0 and $1-\omega_\beta$ observations at 1, where the optimal weights ω_β and $1-\omega_\beta$ are defined in (3.2). From (5.4) the c-efficiency of a design ξ with support points 0 and 1 and weights ω and $1-\omega$ respectively is given by

$$eff_c(\xi) = \omega(1-\omega)/((1-\omega)(\omega_\beta)^2 + \omega(1-\omega_\beta)^2) := u(\omega, \omega_\beta)$$

and the standardised maximin c-optimal criterion is

$$\Phi(\xi) = \min\left\{u(\omega, \omega_\beta) \;\middle|\; \omega_\beta \in [\omega_{\beta_0}, \omega_{\beta_1}]\right\}$$

As in the proof of Theorem 4 it can be shown that for fixed ω the function $\omega_\beta \to u(\omega, \omega_\beta)$ is unimodal and the standardised maximin design ω^* is in $M_=$. Hence we can find ω^* by solving the equation $u(\omega, \omega_{\beta_0}) = u(\omega, \omega_{\omega_1})$ which yields $\omega^* = (\omega_{\beta_0} + \omega_{\beta_1})/2$.

References

Atkinson, A. C., Donev, A. N. and Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS*. Oxford: Oxford University Press.

Bandyopadhyay, U., Biswas, A. and Bhattacharya, R. (2010). A covariate-adjusted adaptive design for two-stage clinical trials with survival data. *Statist. Neerlandica* **64**, 202-226.

Becker, N., McDonald, B. and Khoo, C. (1989). Optimal designs for fitting a proportional hazards regression model to data subject to censoring. *Australian Journal of Statistics* **31**, 449-468.

Biedermann, S., Dette, H. and Zhu, W. (2006). Optimal designs for dose-response models with restricted design spaces. *J. Amer. Statist. Assoc.* **101**, 747-759.

Biedermann, S. and Woods, D. C. (2011). Optimal designs for generalized non-linear models with application to second-harmonic generation experiments. *Applied Statistics* **60**, 281-299.

Statistica Sinica: Preprint
doi:10.5705/ss.2011.271

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* **24**, 586-602.

Dette, H. (1997). Designing experiments with respect to 'standardized' optimality criteria. *J. Roy. Statist. Soc. Ser. B* **59**, 97-110.

Dette, H. and Sahm, M. (1998). Minimax optimal designs in nonlinear regression models. *Statist. Sinica* **8**, 1249-1264.

Dette, H., Melas, V. B. and Wong, W. K. (2006). Locally D-optimal designs for exponential regression models. *Statist. Sinica* **16**, 789-803.

Dette, H. and Melas, V. B. (2011). A note on the de la Garza phenomenon for locally optimal designs. *Ann. Statist.* **39**, 1266-1281.

Dror, H. A. and Steinberg, D. M. (2006). Robust experimental design for multivariate generalized linear models. *Technometrics* **48**, 520-529.

Ford, I., Torsney, B. and Wu, C. F. J. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *J. Roy. Statist. Soc. Ser. B* **54**, 569-583.

Freireich, E. J., Gehan, E., Frei III, E., Schroeder, L. R., Wolman, I. J., Anbari, R., Burgert, O. E., Mills, S. D., Pinkel, D., Selawry, O. S., Moon, J. H., Gendel, B. R., Spurr, C. L., Storrs, R., Haurani, F., Hoogstraten, B. and Lee, S. (1963). The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukimia: a model for evaluation of other potentially useful therapy. *Blood* **21**, 699-716.

Hedayat, A. S., Zhong, J. and Nie, L. (2004). Optimal and efficient designs for 2-parameter nonlinear models. *J. Statist. Plann. Inference* **124**, 205-217.

López-Fidalgo, J., Rivas-López, M. J. and Del Campo, R. (2009). Optimal designs for Cox regression. *Statist. Neerlandica* **63**, 135-148.

McGree, J. M. and Eccleston, J. A. (2010). Investigating design for survival models. *Metrika* **72**, 295-311.

Pukelsheim, F. and Rieder, S. (1992). Efficient rounding of approximate designs. *Biometrika* **79**, 763-770.

doi:10.5705/ss.2011.271

Pukelsheim, F. and Torsney, B. (1991). Optimal weights for experimental designs on linearly independent support points. *Ann. Statist.* **19**, 1614-1625.

Sebastiani, P. and Settimi, R. (1997). A note on D-optimal designs for a logistic regression model. *J. Statist. Plann. Inference* **59**, 359-368.

Silvey, S. D. (1980). *Optimal Designs: An Introduction to the Theory for Parameter Estimation.* Chapman and Hall, London.

Sitter, R. R. and Torsney, B. (1992). D-optimal designs for generalized linear models. Technical Report No. 211, Technical Report Series of the Laboratory for Research in Statistics and Probability, Carleton University, Ottawa, Canada.

Sitter, R. R. and Torsney, B. (1995). Optimal designs for binary response experiments with two design variables, *Statist. Sinica* **5**, 405-419.

Woods, D. C., Lewis, S. M., Eccleston, J. A. and Russell, K. G. (2006). Designs for generalized linear models with several variables and model uncertainty. *Technometrics* **48**, 284-292.

Wu, S. J., Lin, Y. P. and Chen, Y. J. (2006). Planning step-stress life test with progressively type I group-censored exponential data. *Statist. Neerlandica* **60**, 46-56.

Xu, X. (2009). Robust prediction and extrapolation designs for censored data. *J. Statist. Plann. Inference* **139**, 486-502.

Yang, M. and Stufken, J. (2009). Support points of locally optimal designs for nonlinear models with two parameters. *Ann. Statist.* **37**, 518-541.

Yang, M. (2010). On the de la Garza phenomenon. *Ann. Statist.* **38**, 2499-2524.

Southampton Statistical Sciences Research Institute, University of Southampton, University Road, Southampton SO17 1BJ, United Kingdom

E-mail: mk21g09@southampton.ac.uk

E-mail: S.Biedermann@southampton.ac.uk

E-mail: A.C.Kimber@southampton.ac.uk